Case study

「合理的創薬」に必須なタンパク質の理解に向けて 研究者のための情報処理インフラ構築に挑む

既知の実験データからタンパク質全体の立体構造をモデル化する一連のシステムを HP ProLiant DL980 G7上で統合し、創薬をサポートできる研究基盤を提供する

業界
研究機関

目的
創薬研究を支援する情報処理基盤の構築と運用

アプローチ
- 多様な複数の処理を柔軟に実行するために、1筐体で完結するマルチプロセッサーサーバーを活用
- コストパフォーマンスを評価し、高い処理性能と優れた信頼性、拡張性を備えたインテル® Xeon® プロセッサー E7ファミリー搭載の8ソケットサーバー、HP ProLiant DL980 G7を導入
- データベースとの連携も考慮し、高速I/Oと高い信頼性を両立したSSD、HP VMAシリーズを外付けストレージに採用

ITの効果
- 処理内容に応じて、柔軟なリソース配分が可能
- テスト運用でも確実にプログラムの処理スピードが向上
- クラスター環境と比べ、日常的な運用管理は手間要らずに

ビジネスの効果
- プログラムなどを最適化することで、モデリング時間のさらなる短縮が可能に
- タンパク質全体の立体モデルが容易に得られることで、創薬の取り組みが加速

タンパク質の立体構造モデリング

いくつかサーバー候補があった中で、これだけのコア数を1つの筐体内に収めることができ、コストパフォーマンスにも優れていたのが HP ProLiant DL980 G7でした。強力な演算パワーが必要な時は複数のコアで高速に並列処理が行え、異なる複数の処理をコアごとに割り振ることもできる。状況に応じて使い方を柔軟に切り替えたいと考えていましたから、シングルプロセッサーを集めたクラスターではなく、一つの筐体で済むマルチプロセッサーサーバーが大前提でした

白井 剛氏　長浜バイオ大学 バイオサイエンス学部 教授

長浜バイオ大学はバイオサイエンスに関する最先端の研究と教育を行っている、国内でもユニークな単科大学だ。特に、バイオサイエンスとコンピューター技術の両方を融合させた研究と人材育成に取り組んでいるコンピュータバイオサイエンス学科は、この領域では日本で最初に開設された学科であり、一つの建物内にバイオインフォマティック関連の様々な研究者が集まって活発な共同研究や情報交換を行っている。こうした特長を活かし、同学科では、体内にあるタンパク質が機能する時に形成される超分子の立体構造をモデル化するための研究プロジェクトが動き出している。「創薬」という医薬品開発の新たなアプローチをコンピューター技術でサポートしようというものだ。

長浜バイオ大学
バイオサイエンス学部
教授
白井 剛 氏

長浜バイオ大学
バイオサイエンス学部
講師
塩生 真史 氏

タンパク質の立体構造モデリング

タンパク質の挙動と構造から医薬品を設計する「合理的創薬」

病気の治療などに使われる医薬品の開発は、従来、化学的な合成、動植物からの抽出といった方法で得た物質を分析・研究し、薬効のある成分を探し当てるという手法で行われてきた。しかし、近年、人の遺伝子情報の解読が急速に進展、さらにはこの遺伝子情報を基に体内で作り出され、あらゆる生命活動を支えている膨大な数のタンパク質などに関して、新しい知見を得ようとする「ポストゲノム」研究が凄まじい勢いで展開されている。こうした取り組みによって、これまでとまったく異なるアプローチで医薬品を開発することが可能になってきた。

新たな手法は、遺伝子やタンパク質などに関する知見を基に、病気の原因となる、あるいは病気に関係しているタンパク質を突き止め、これに直接作用する医薬品を研究者が理論的にデザインすることから、「合理的創薬」と呼ばれている。従来の医薬品と比べ、より効果が高く、副作用はより小さい医薬品を開発できると考えられており、医療分野からの期待も非常に大きい。

国も創薬の取り組みを積極的に支援しようと、様々な補助事業を実施。文部科学省では、創薬や医療技術の研究をサポートする基盤の強化を目的に「創薬等ライフサイエンス研究支援基盤事業」を2012年度からスタートさせた。この補助事業の一つとして、創薬に重要な役割を持つタンパク質の構造解析などを加速すべく、研究に不可欠な情報処理インフラなどの整備を促すための「創薬等支援技術基盤プラットフォーム事業」を実施している。

バイオサイエンスに特化したユニークな教育で知られる長浜バイオ大学では、バイオサイエンス学部コンピュータバイオサイエンス学科の白井 剛教授を中心としたグループが、この事業にデータベース技術やコンピューターシミュレーション技術などを統合した情報処理基盤の構築プロジェクトを提案し、採択された。5年にも及ぶ息の長いこのプロジェクトの研究インフラとして、高い処理性能と優れた信頼性、拡張性を備えたインテル® Xeon® プロセッサー E7ファミリー搭載のマルチプロセッサーサーバー、HP ProLiant DL980 G7を中心としたシステムが稼働を開始している。

タンパク質の構成パーツから全体の立体構造をモデル化する

白井教授を含め、同大学のコンピュータバイオサイエンス学科に所属する研究者6名で構成されるグループが構想したプロジェクトのテーマは「超分子モデリングパイプラインの構築」というものである。

タンパク質が何らかの機能を発揮しようとする際、生体内ではタンパク質や核酸といった分子がゆるやかに結合した巨大な超分子が一時的に形成されるという現象が知られている。合理的創薬に取り組むには、超分子の働くメカニズムや反応に関与している超分子の部位などを明らかにする必要があり、それには超分子の3次元的な立体構造を解明することが前提になってくる。

「ただし、超分子全体の立体構造を実験データから解明することは、現状では極めて困難です。超分子は非常に巨大であるため、実験技術的な制約から、全体構造を決定するための実験データを直接、かつ包括的に得ることができないのです。その一方で、超分子のパーツともいえる分子量の小さなタンパク質は、正確な立体構造データが得られます。ここ数年、立体構造データを世界規模で集積するデータベース『プロテインデータバンク（PDB）』に登録される数は、指数関数的な勢いで増えています」と、白井教授はプロジェクトの背景を解説する。

そこで白井教授たちは、PDBなどに集積されている既知の立体構造データを活用し、ドッキングシミュレーション、粗視化分子動力学計算といったいくつかの最新バイオインフォマティック技術を連携させた情報処理のパイプラインを構築することで、超分子全体の立体構造をモデリングするというアイディアにたどり着いた。「創薬に関わる様々な分野の研究者が支援を得たいと考えた時に、プログラムなどの基盤技術が存在して、これをオペレートできる研究者がいる、という環境を作ることがプロジェクトの最終目標です」（白井教授）。

これまでの研究成果を連携し一連のパイプラインとして統合する

白井教授たちが考案したパイプラインは四つのステップで構成される。まず第1ステップは、パーツとなるタンパク質のデータが膨大に蓄積された公開データベースからモデリングに必要な情報を抽出し、分類・整理して知識ベースを構築。この情報を基に簡易な手法でパーツを組み合わせ、超分子を大まかにモデリングする。第2のステップは、ドメイン・サブユニットドッキングというより高度なモデリング手法を用い、知識ベースの情報も参照しながら超分子をモデリングする。そして第3ステップでは、第1および第2のステップで作成したモデルを、分子動力学的な計算を用いて、より安定した構造になるよう調整。最終ステップでは、得られたモデルデータを創薬に関わる研究者たちに利用しやすい形で表示させる。

タンパク質の立体構造モデリング

■パイプラインのハードウェア環境

**インテル® Xeon® プロセッサー
E7ファミリー**

HP VMA3205（5TB）

ハードウェア:HP ProLiant DL980 G7
搭載プロセッサー:インテル® Xeon® プロセッサー E7-4870 2.4GHz 8基／80コア
搭載メモリー:1TB
搭載記憶領域:900GB SAS HDD×8台　合計7.2TB
OS:CentOS

「今回のプロジェクトで取り組むべきテーマは、大きく二つあります。まず、パイプラインを実現するためのプログラムやその背景となるアルゴリズムなどを開発し、これらをマッシュアップして一連のパイプラインとして機能する情報処理基盤にまとめること。そして、この情報処理基盤を使って、創薬に関わる外部の研究者から寄せられる要請に対応し、研究をサポートすること。前者の取り組みを『高度化』、後者を『支援』と呼んでいますが、この二つを達成するには、非常に多様な要求に応えてくれるサーバー環境が必要でした」（白井教授）。

たとえば、第1ステップではデータベースからのデータ読み出し、処理結果の書き出しなどを膨大に繰り返す。このためファイルI/Oのスピードが極めて重要になる。第2ステップや第3ステップでは演算のスピードがモデリングなどの処理にかかる時間を左右する。「併せて、『支援』の面では、Webベースでのサービス提供を見込んでいるものの、利用者からのリクエストに応じて短時間に結果を表示することを目指しています。そのためにはモデリングや分子動力学計算を高速で実行する必要があり、演算パワーがやはり不可欠です。運用時の管理性も考慮して、これらを一つの環境で実現したいと考えていたのです」と白井教授は語る。

1筐体にマルチコアを格納でき 管理も容易なHP ProLiant DL980 G7

5年にわたるプロジェクトを支えるサーバー環境として、白井教授たちが選択したのは8CPU/80コアを搭載したHP ProLiant DL980 G7だった。

「いくつかサーバー候補があった中で、これだけのコア数を一つの筐体内に収めることができ、コストパフォーマンスにも優れていたのがHP ProLiant DL980 G7でした」と白井教授。「強力な演算パワーが必要な時は複数のコアで高速に並列処理が行え、異なる複数の処理をコアごとに割り振ることもできる。状況に応じて使い方を柔軟に切り替えたいと考えていましたから、シングルプロセッサーを集めたクラスターではなく、一つの筐体で済むマルチプロセッサーサーバーが大前提でした」。

管理の面でも、複数のサーバーを組み合わせるクラスターと比べ、マルチプロセッサーサーバーには大きなメリットがあった。「私の担当する第1ステップでは、特にファイルの読み書きを延々と繰り返すため、クラスター環境ではネットワークを占有してしまう時があり、同僚の研究者から苦情が来ることがありました。しかしHP ProLiant DL980 G7ならネットワークの問題を考慮しなくて済みます。また、複数のプログラムをクラスター環境で運用する場合、異なるファイルシステムをどのように扱うかが悩みのタネでした。こうした対応も簡単に実現できると思いました」（白井教授）。

さらに、高速なファイルI/Oを実現できるSSDストレージがオプションとして用意されていたこともHP ProLiant DL980 G7の大きな魅力だったようだ。5年にわたって長期間使っていくことになるため、信頼性も重要だ。白井教授たちが外部ストレージとして採用したHP VMAシリーズは、信頼性の高いSLC（Single Level Cell）NANDフラッシュメモリーを搭載しており、1TB/時の書き込みを行うヘビーな用途でも5年は問題なく使用できるというアドバンテージを備えている。フラッシュテクノロジーには可動部品が含まれないこともあり、ハードディスクと比べ、トータルな信頼性は段違いだ。

まだ試運転の段階にあるものの 各ステップの処理時間は確実に短縮

2012年9月末、HP ProLiant DL980 G7を中心としたハードウェア類が納入され、パイプライン構築に向けた作業が動き出した。これまでのところ、従来、開発してきた各種プログラムをHP

ProLiant DL980 G7上で走らせ、処理の高速化を図るために必要な開発ポイントを見極める作業が進んでいる。運用は緒に就いたばかりだが、HP ProLiant DL980 G7を選択したことのメリットを確認できる初期の成果が徐々に得られているようだ。

第2ステップを担当するコンピュータバイオサイエンス学科の塩生真史講師は、ドッキング処理を高速化できたという。「まだまだ試運転という段階で、これまでワークステーションで走らせていたプログラムをそのまま使っていますが、36コアを使って演算時間を比較してみたところ、およそ7倍のスピードアップが図れています。今後は、アルゴリズムやプログラムを根本から見直し、さらなるスピードアップを目指すつもりです」。

塩生講師はサーバーの運用も担当しており、運用管理の負担が減ったことも喜んでいる。「ユーザーを増やす、共用するアプリケーションをインストールする、といった単純な作業でも、クラスター環境では複数の筐体に対して作業を行う場合があり、意外と手間がかかりました。1筐体内ですべてのコアを制御しているHP ProLiant DL980 G7の環境に変わったことで操作が一度で済み、本当に楽になりました」。

白井教授も手ごたえを感じている。「知識ベースの構築に要する時間が半分になっています。以前とほぼ同じコア数で作業をしているので、これはHP ProLiant DL980 G7のI/O性能が純粋に効いている結果だと思います」。

膨大なリソースをフル活用するには発想を大きく転換する必要がある

プロジェクトとしての当面の目標は、パイプラインの各ステップを高速化し、パイプライン全体としての処理時間を、Webで利用してもそれほど違和感のない程度までできる限り短縮していくこと。こうした作業を今後2年をメドに達成したいとチームでは考えている。

「正直なところ、総コア数80、メモリー搭載量1TBという、これまでは想像もできなかったHP ProLiant DL980 G7の膨大なリソースに戸惑っている、といのが現状の感想です。これまで使ってきたワークステーションとはまったく発想を変えて、アルゴリズムやプログラムを作り替えていく必要があるでしょう。しかし、HP ProLiant DL980 G7に最適化した処理系を構築できれば、パイプライン全体の処理時間は劇的に短縮できるはず。これからが楽しみです」と白井教授は期待を込める。

超分子全体の立体構造をモデル化し、「合理的創薬」というアプローチで新たな医薬品を生み出すという世界でも最先端の試みに、HP ProLiant DL980 G7の持つポテンシャルが役立つ日は、確実に近づいている。

ソリューション概略

導入ハードウェア
HP ProLiant DL980 G7
HP VMA3205 5TB SLCメモリーアレイ

お問い合わせはカスタマー・インフォメーションセンターへ
03-5749-8328 月～金 9:00～19:00　土 10:00～17:00（日、祝祭日、年末年始および5/1を除く）
機器のお見積もりについては、代理店、または弊社営業にご相談ください。
HP ProLiantに関する情報は http://www.hp.com/jp/proliant

本カタログは、環境に配慮した用紙と植物性大豆油インキを使用しています。
日本ヒューレット・パッカード株式会社
〒136-8711 東京都江東区大島2-2-1